

Scaling Language Models: Methods, Analysis & Insights from Training Gopher

Year: 2021 | Citations: 1484 | Authors: Jack W. Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann

Abstract

Language modelling provides a step towards intelligent communication systems by harnessing large repositories of written human knowledge to better predict and understand the world. In this paper, we present an analysis of Transformer-based language model performance across a wide range of model scales -- from models with tens of millions of parameters up to a 280 billion parameter model called Gopher. These models are evaluated on 152 diverse tasks, achieving state-of-the-art performance across the majority. Gains from scale are largest in areas such as reading comprehension, fact-checking, and the identification of toxic language, but logical and mathematical reasoning see less benefit. We provide a holistic analysis of the training dataset and model's behaviour, covering the intersection of model scale with bias and toxicity. Finally we discuss the application of language models to AI safety and the mitigation of downstream harms.